

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANTS: ARIYASU, Toshio
NAKAMURA, Shuji
ORITA, Kunzo

(ii) TITLE OF INVENTION: HEDGEHOG PROTEIN

(iii) NUMBER OF SEQUENCES: 31

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: BROWDY AND NEIMARK
(B) STREET: 419 Seventh Street N.W., Ste. 300
(C) CITY: Washington
(D) STATE: D.C.
(E) COUNTRY: United States of America
(F) ZIP: 20004

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER: 09/063,778
(B) FILING DATE: 22-APR-1998

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: JP 97-121578
(B) FILING DATE: 25-APR-1997

(viii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: JP 98-
(B) FILING DATE: 14-APR-1998

(ix) ATTORNEY/AGENT INFORMATION:

(A) NAME: Browdy, Roger L.
(B) REGISTRATION NUMBER: 25,618
(C) REFERENCE/DOCKET NUMBER: ARIYASU-1

(x) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: (202) 628-5197
(B) TELEFAX: (202) 737-35281

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 176 amino acids
(B) TYPE: amino acid
(C) strandedness: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

Cys Gly Pro Gly Arg Gly Pro Val Gly Arg Arg Arg Tyr Ala Arg Lys
1 5 10 15

Gln Leu Val Pro Leu Leu Tyr Lys Gln Phe Val Pro Gly Val Pro Glu
20 25 30

Arg Thr Leu Gly Ala Ser Gly Pro Ala Glu Gly Arg Val Ala Arg Gly
35 40 45

RECEIVED
98 JUN 11 AM 7:23
GROUP 2700

Ser Glu Arg Phe Arg Asp Leu Val Pro Asn Tyr Asn Pro Asp Ile Ile
 50 55 60
 Phe Lys Asp Glu Glu Asn Ser Gly Ala Asp Arg Leu Met Thr Glu Arg
 65 70 75 80
 Cys Lys Glu Arg Val Asn Ala Leu Ala Ile Ala Val Met Asn Met Trp
 85 90 95
 Pro Gly Val Arg Leu Arg Val Thr Glu Gly Trp Asp Glu Asp Gly His
 100 105 110
 His Ala Gln Asp Ser Leu His Tyr Glu Gly Arg Ala Leu Asp Ile Thr
 115 120 125
 Thr Ser Asp Arg Asp Arg Asn Lys Tyr Gly Leu Leu Ala Arg Leu Ala
 130 135 140
 Val Glu Ala Gly Phe Asp Trp Val Tyr Tyr Glu Ser Arg Asn His Ile
 145 150 155 160
 His Val Ser Val Lys Ala Asp Asn Ser Leu Ala Val Arg Ala Gly Gly
 165 170 175

(2) INFORMATION FOR SEQ ID NO: 2:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 374 amino acids
 - (B) TYPE: amino acid
 - (C) strandedness: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (ix) FEATURE:
 - (A) NAME/KEY: mat peptide
 - (B) LOCATION: 1..176
 - (C) IDENTIFICATION METHOD: S
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Cys Gly Pro Gly Arg Gly Pro Val Gly Arg Arg Arg Tyr Ala Arg Lys
 1 5 10 15
 Gln Leu Val Pro Leu Leu Tyr Lys Gln Phe Val Pro Gly Val Pro Glu
 20 25 30
 Arg Thr Leu Gly Ala Ser Gly Pro Ala Glu Gly Arg Val Ala Arg Gly
 35 40 45
 Ser Glu Arg Phe Arg Asp Leu Val Pro Asn Tyr Asn Pro Asp Ile Ile
 50 55 60
 Phe Lys Asp Glu Glu Asn Ser Gly Ala Asp Arg Leu Met Thr Glu Arg
 65 70 75 80
 Cys Lys Glu Arg Val Asn Ala Leu Ala Ile Ala Val Met Asn Met Trp
 85 90 95
 Pro Gly Val Arg Leu Arg Val Thr Glu Gly Trp Asp Glu Asp Gly His
 100 105 110
 His Ala Gln Asp Ser Leu His Tyr Glu Gly Arg Ala Leu Asp Ile Thr
 115 120 125
 Thr Ser Asp Arg Asp Arg Asn Lys Tyr Gly Leu Leu Ala Arg Leu Ala
 130 135 140

Val Glu Ala Gly Phe Asp Trp Val Tyr Tyr Glu Ser Arg Asn His Ile
 145 150 155 160
 His Val Ser Val Lys Ala Asp Asn Ser Leu Ala Val Arg Ala Gly Gly
 165 170 175
 Cys Phe Pro Gly Asn Ala Thr Val Arg Leu Trp Ser Gly Glu Arg Lys
 180 185 190
 Gly Leu Arg Glu Leu His Arg Gly Asp Trp Val Leu Thr Ala Asp Ala
 195 200 205
 Ser Gly Arg Val Val Pro Thr Pro Val Leu Leu Phe Leu Asp Arg Asp
 210 215 220
 Leu Gln Arg Arg Ala Ser Phe Val Ala Val Glu Thr Glu Trp Pro Pro
 225 230 235 240
 Arg Lys Leu Leu Leu Thr Pro Trp His Leu Val Phe Ala Ala Arg Gly
 245 250 255
 Pro Ala Pro Ala Pro Gly Asp Phe Ala Pro Val Phe Ala Arg Arg Leu
 260 265 270
 Arg Ala Gly Asp Ser Val Leu Ala Pro Gly Gly Asp Ala Leu Arg Pro
 275 280 285
 Ala Arg Val Ala Arg Val Ala Arg Glu Glu Ala Val Gly Val Phe Ala
 290 295 300
 Pro Leu Thr Ala His Gly Thr Leu Leu Val Asn Asp Val Leu Ala Ser
 305 310 315 320
 Cys Tyr Ala Val Leu Glu Ser His Gln Trp Ala His Arg Ala Phe Ala
 325 330 335
 Pro Leu Arg Leu Leu His Ala Leu Gly Ala Leu Leu Pro Gly Gly Ala
 340 345 350
 Val Gln Pro Thr Gly Met His Trp Tyr Ser Arg Leu Leu Tyr Arg Leu
 355 360 365
 Ala Glu Glu Leu Leu Gly
 370

(2) INFORMATION FOR SEQ ID NO: 3:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 396 amino acids
 - (B) TYPE: amino acid
 - (C) strandedness: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (ix) FEATURE:
 - (A) NAME/KEY: sig peptide
 - (B) LOCATION: -22...-1
 - (C) IDENTIFICATION METHOD: S
 - (A) NAME/KEY: mat peptide
 - (B) LOCATION: 1..176
 - (C) IDENTIFICATION METHOD: S

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

Met Ala Leu Leu Thr Asn Leu Leu Pro Leu Cys Cys Leu Ala Leu Leu
 -20 -15 -10

Ala Leu Pro Ala Gln Ser Cys Gly Pro Gly Arg Gly Pro Val Gly Arg
-5 1 5 10

Arg Arg Tyr Ala Arg Lys Gln Leu Val Pro Leu Leu Tyr Lys Gln Phe
15 20 25

Val Pro Gly Val Pro Glu Arg Thr Leu Gly Ala Ser Gly Pro Ala Glu
30 35 40

Gly Arg Val Ala Arg Gly Ser Glu Arg Phe Arg Asp Leu Val Pro Asn
45 50 55

Tyr Asn Pro Asp Ile Ile Phe Lys Asp Glu Glu Asn Ser Gly Ala Asp
60 65 70

Arg Leu Met Thr Glu Arg Cys Lys Glu Arg Val Asn Ala Leu Ala Ile
75 80 85 90

Ala Val Met Asn Met Trp Pro Gly Val Arg Leu Arg Val Thr Glu Gly
95 100 105

Trp Asp Glu Asp Gly His His Ala Gln Asp Ser Leu His Tyr Glu Gly
110 115 120

Arg Ala Leu Asp Ile Thr Thr Ser Asp Arg Asp Arg Asn Lys Tyr Gly
125 130 135

Leu Leu Ala Arg Leu Ala Val Glu Ala Gly Phe Asp Trp Val Tyr Tyr
140 145 150

Glu Ser Arg Asn His Ile His Val Ser Val Lys Ala Asp Asn Ser Leu
155 160 165 170

Ala Val Arg Ala Gly Gly Cys Phe Pro Gly Asn Ala Thr Val Arg Leu
175 180 185

Trp Ser Gly Glu Arg Lys Gly Leu Arg Glu Leu His Arg Gly Asp Trp
190 195 200

Val Leu Thr Ala Asp Ala Ser Gly Arg Val Val Pro Thr Pro Val Leu
205 210 215

Leu Phe Leu Asp Arg Asp Leu Gln Arg Arg Ala Ser Phe Val Ala Val
220 225 230

Glu Thr Glu Trp Pro Pro Arg Lys Leu Leu Leu Thr Pro Trp His Leu
235 240 245 250

Val Phe Ala Ala Arg Gly Pro Ala Pro Ala Pro Gly Asp Phe Ala Pro
255 260 265

Val Phe Ala Arg Arg Leu Arg Ala Gly Asp Ser Val Leu Ala Pro Gly
270 275 280

Gly Asp Ala Leu Arg Pro Ala Arg Val Ala Arg Val Ala Arg Glu Glu
285 290 295

Ala Val Gly Val Phe Ala Pro Leu Thr Ala His Gly Thr Leu Leu Val
300 305 310

Asn Asp Val Leu Ala Ser Cys Tyr Ala Val Leu Glu Ser His Gln Trp
315 320 325 330

Ala His Arg Ala Phe Ala Pro Leu Arg Leu Leu His Ala Leu Gly Ala
335 340 345

Leu Leu Pro Gly Gly Ala Val Gln Pro Thr Gly Met His Trp Tyr Ser
350 355 360

Arg Leu Leu Tyr Arg Leu Ala Glu Glu Leu Leu Gly
365 370

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 528 base pairs
- (B) TYPE: nucleic acid
- (C) strandedness: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

- (A) NAME/KEY: mat peptide
- (B) LOCATION: 1..528
- (C) IDENTIFICATION METHOD: S

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

TGC	GGG	CCG	GGC	CGG	GGG	CCG	GTT	GGC	CGG	CGC	CAC	TAT	GCG	CGC	AAG	48
Cys	Gly	Pro	Gly	Arg	Gly	Pro	Val	Gly	Arg	Arg	Arg	Tyr	Ala	Arg	Lys	
1				5				10					15			
CAG	CTC	GTG	CCG	CTA	CTC	TAC	AAG	CAA	TTT	GTG	CCC	GGC	GTG	CCA	GAG	96
Gln	Leu	Val	Pro	Leu	Leu	Tyr	Lys	Gln	Phe	Val	Pro	Gly	Val	Pro	Glu	
				20				25					30			
CGG	ACC	CTG	GGC	GCC	AGT	GGG	CCA	GCG	GAG	GGG	AGG	GTG	GCA	AGG	GGC	144
Arg	Thr	Leu	Gly	Ala	Ser	Gly	Pro	Ala	Glu	Gly	Arg	Val	Ala	Arg	Gly	
	35				40			45								
TCC	GAG	CGC	TTC	CGG	GAC	CTC	GTG	CCC	AAC	TAC	AAC	CCC	GAC	ATC	ATC	192
Ser	Glu	Arg	Phe	Arg	Asp	Leu	Val	Pro	Asn	Tyr	Asn	Pro	Asp	Ile	Ile	
	50			55			60									
TTC	AAG	GAT	GAG	GAG	AAC	AGT	GGA	GCC	GAC	CGC	CTG	ATG	ACC	GAA	CGT	240
Phe	Lys	Asp	Glu	Glu	Asn	Ser	Gly	Ala	Asp	Arg	Leu	Met	Thr	Glu	Arg	
	65			70			75						80			
TGT	AAG	GAA	CGG	GTG	AAC	GCT	TTG	GCC	ATT	GCC	GTG	ATG	AAC	ATG	TGG	288
Cys	Lys	Glu	Arg	Val	Asn	Ala	Leu	Ala	Ile	Ala	Val	Met	Asn	Met	Trp	
	85				90								95			
CCC	GGA	GTG	CGC	CTA	CGA	GTG	ACT	GAG	GGC	TGG	GAC	GAG	GAC	GGC	CAC	336
Pro	Gly	Val	Arg	Leu	Arg	Val	Thr	Glu	Gly	Trp	Asp	Glu	Asp	Gly	His	
	100				105								110			
CAC	GCT	CAG	GAT	TCA	CTC	CAC	TAC	GAA	GGC	CGT	GCT	TTG	GAC	ATC	ACT	384
His	Ala	Gln	Asp	Ser	Leu	His	Tyr	Glu	Gly	Arg	Ala	Leu	Asp	Ile	Thr	
	115				120								125			
ACG	TCT	GAC	CGC	GAC	CGC	AAC	AAG	TAT	GGG	TTG	CTG	GCG	CGC	CTC	GCA	432
Thr	Ser	Asp	Arg	Asp	Arg	Asn	Lys	Tyr	Gly	Leu	Leu	Ala	Arg	Leu	Ala	
	130			135									140			
GTG	GAA	GCC	GGC	TTC	GAC	TGG	GTC	TAC	TAC	GAG	TCC	CGC	AAC	CAC	ATC	480
Val	Glu	Ala	Gly	Phe	Asp	Trp	Val	Tyr	Tyr	Glu	Ser	Arg	Asn	His	Ile	
	145			150							155			160		
CAC	GTG	TCG	GTC	AAA	GCT	GAT	AAC	TCA	CTG	GCG	GTC	CGG	GCG	GGC	GGC	528
His	Val	Ser	Val	Lys	Ala	Asp	Asn	Ser	Leu	Ala	Val	Arg	Ala	Gly	Gly	
	165				170								175			

(2) INFORMATION FOR SEQ ID NO: 5:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1122 base pairs
 - (B) TYPE: nucleic acid
 - (C) strandedness: double
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (ix) FEATURE:
 - (A) NAME/KEY: mat peptide
 - (B) LOCATION: 1..528
 - (C) IDENTIFICATION METHOD: S

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

TGC	GGG	CCG	GGC	CGG	GGG	CCG	GTT	GGC	CGG	CGC	CGC	TAT	GCG	CGC	AAG	48
Cys	Gly	Pro	Gly	Arg	Gly	Pro	Val	Gly	Arg	Arg	Arg	Tyr	Ala	Arg	Lys	
1									10						15	
CAG	CTC	GTG	CCG	CTA	CTC	TAC	AAG	CAA	TTT	GTG	CCC	GGC	GTG	CCA	GAG	96
Gln	Leu	Val	Pro	Leu	Leu	Tyr	Lys	Gln	Phe	Val	Pro	Gly	Val	Pro	Glu	
									25						30	
CGG	ACC	CTG	GGC	GCC	AGT	GGG	CCA	GCG	GAG	GGG	AGG	GTG	GCA	AGG	GGC	144
Arg	Thr	Leu	Gly	Ala	Ser	Gly	Pro	Ala	Glu	Gly	Arg	Val	Ala	Arg	Gly	
									40						45	
TCC	GAG	CGC	TTC	CGG	GAC	CTC	GTG	CCC	AAC	TAC	AAC	CCC	GAC	ATC	ATC	192
Ser	Glu	Arg	Phe	Arg	Asp	Leu	Val	Pro	Asn	Tyr	Asn	Pro	Asp	Ile	Ile	
									55						60	
TTC	AAG	GAT	GAG	GAG	AAC	AGT	GGG	GCC	GAC	CGC	CTG	ATG	ACC	GAA	CGT	240
Phe	Lys	Asp	Glu	Glu	Asn	Ser	Gly	Ala	Asp	Arg	Leu	Met	Thr	Glu	Arg	
									70						80	
TGT	AAG	GAA	CGG	GTG	AAC	GCT	TTG	GCC	ATT	GCC	GTG	ATG	AAC	ATG	TGG	288
Cys	Lys	Glu	Arg	Val	Asn	Ala	Leu	Ala	Ile	Ala	Val	Met	Asn	Met	Trp	
									85						95	
CCC	GGA	GTG	CGC	CTA	CGA	GTG	ACT	GAG	GGC	TGG	GAC	GAG	GAC	GGC	CAC	336
Pro	Gly	Val	Arg	Leu	Arg	Val	Thr	Glu	Gly	Trp	Asp	Glu	Asp	Gly	His	
									100						110	
CAC	GCT	CAG	GAT	TCA	CTC	CAC	TAC	GAA	GGC	CGT	GCT	TTG	GAC	ATC	ACT	384
His	Ala	Gln	Asp	Ser	Leu	His	Tyr	Glu	Gly	Arg	Ala	Leu	Asp	Ile	Thr	
									115						125	
ACG	TCT	GAC	CGC	GAC	CGC	AAC	AAG	TAT	GGG	TTG	CTG	GCG	CGC	CTC	GCA	432
Thr	Ser	Asp	Arg	Asp	Arg	Asn	Lys	Tyr	Gly	Leu	Leu	Ala	Arg	Leu	Ala	
									130						140	
GTG	GAA	GCC	GGC	TTC	GAC	TGG	GTC	TAC	TAC	GAG	TCC	CGC	AAC	CAC	ATC	480
Val	Glu	Ala	Gly	Phe	Asp	Trp	Val	Tyr	Tyr	Glu	Ser	Arg	Asn	His	Ile	
									145						155	
CAC	GTG	TCG	GTC	AAA	GCT	GAT	AAC	TCA	CTG	GCG	GTC	CGG	GCG	GGC	GGC	528
His	Val	Ser	Val	Lys	Ala	Asp	Asn	Ser	Leu	Ala	Val	Arg	Ala	Gly	Gly	
									165						175	
TGC	TTT	CCG	GGG	AAT	GCA	ACT	GTG	CGC	CTG	TGG	AGC	GGC	GAG	CGG	AAA	576
Cys	Phe	Pro	Gly	Asn	Ala	Thr	Val	Arg	Leu	Trp	Ser	Gly	Glu	Arg	Lys	
									180						190	
GGG	CTG	CGG	GAA	CTG	CAC	CGC	GGG	GAC	TGG	GTT	TTG	ACG	GCC	GAT	GCG	624
Gly	Leu	Arg	Glu	Leu	His	Arg	Gly	Asp	Trp	Val	Leu	Thr	Ala	Asp	Ala	
									195						205	

TCA	GGC	CGG	GTG	GTG	CCC	ACG	CCG	GTG	CTG	CTC	TTC	CTG	GAC	CGG	GAC	672
Ser	Gly	Arg	Val	Val	Pro	Thr	Pro	Val	Leu	Leu	Phe	Leu	Asp	Arg	Asp	
210															220	
TTG	CAG	CGC	CGG	GCT	TCA	TTT	GTG	GCT	GTG	GAG	ACC	GAG	TGG	CCT	CCA	720
Leu	Gln	Arg	Arg	Ala	Ser	Phe	Val	Ala	Val	Glu	Thr	Glu	Trp	Pro	Pro	
225															240	
CGC	AAA	CTG	TTG	CTC	ACG	CCC	TGG	CAC	CTG	GTG	TTT	GCC	GCT	CGA	GGG	768
Arg	Lys	Leu	Leu	Leu	Thr	Pro	Trp	His	Leu	Val	Phe	Ala	Ala	Arg	Gly	
245															255	
CCG	GCG	CCC	GCG	CCA	GGC	GAC	TTT	GCA	CCG	GTG	TTC	GCG	CGC	CGG	CTA	816
Pro	Ala	Pro	Ala	Pro	Gly	Asp	Phe	Ala	Pro	Val	Phe	Ala	Arg	Arg	Leu	
260															270	
CGC	GCT	GGG	GAC	TCG	GTG	CTG	GCG	CCC	GGC	GGG	GAT	GCG	CTT	CGG	CCA	864
Arg	Ala	Gly	Asp	Ser	Val	Leu	Ala	Pro	Gly	Gly	Asp	Ala	Leu	Arg	Pro	
275															285	
GGC	CGC	GTG	GCC	CGT	GTG	GCG	CGG	GAG	GAA	GCC	GTG	GGC	GTG	TTC	GCG	912
Ala	Arg	Val	Ala	Arg	Val	Ala	Arg	Glu	Glu	Ala	Val	Gly	Val	Phe	Ala	
290															300	
CCG	CTC	ACC	GCG	CAC	GGG	ACG	CTG	CTG	GTG	AAC	GAT	GTC	CTG	GCC	TCT	960
Pro	Leu	Thr	Ala	His	Gly	Thr	Leu	Leu	Val	Asn	Asp	Val	Leu	Ala	Ser	
305															320	
TGC	TAC	GCG	GTT	CTG	GAG	AGT	CAC	CAG	TGG	GCG	CAC	CGC	GCT	TTT	GCC	1008
Cys	Tyr	Ala	Val	Leu	Glu	Ser	His	Gln	Trp	Ala	His	Arg	Ala	Phe	Ala	
325															335	
CCC	TTG	AGA	CTG	CTG	CAC	GCG	CTA	GGG	GCG	CTG	CTC	CCC	GGC	GGG	GCC	1056
Pro	Leu	Arg	Leu	Leu	His	Ala	Leu	Gly	Ala	Leu	Leu	Pro	Gly	Gly	Ala	
340															350	
GTC	CAG	CCG	ACT	GGC	ATG	CAT	TGG	TAC	TCT	CGG	CTC	CTC	TAC	CGC	TTA	1104
Val	Gln	Pro	Thr	Gly	Met	His	Trp	Tyr	Ser	Arg	Leu	Leu	Tyr	Arg	Leu	
355															365	
GGC	GAG	GAG	CTA	CTG	GGC											1122
Ala	Glu	Glu	Leu	Leu	Gly											
370																

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1188 base pairs
- (B) TYPE: nucleic acid
- (C) strandedness: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

- (A) NAME/KEY: sig peptide
- (B) LOCATION: 1..66
- (C) IDENTIFICATION METHOD: S
- (A) NAME/KEY: mat peptide
- (B) LOCATION: 67..594
- (C) IDENTIFICATION METHOD: S

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

ATG	GCT	CTC	CTG	ACC	AAT	CTA	CTG	CCC	CTG	TGC	TGC	TTG	GCA	CTT	CTG	48
Met	Ala	Leu	Leu	Thr	Asn	Leu	Leu	Pro	Leu	Cys	Cys	Leu	Ala	Leu	Leu	
-20															-10	

GCG CTG CCA GCC CAG AGC TGC GGG CCG GGC CGG GGG CCG GTT GGC CGG	96
Ala Leu Pro Ala Gln Ser Cys Gly Pro Gly Arg Gly Pro Val Gly Arg	
-5 1 5 10	
CGC CGC TAT GCG CGC AAG CAG CTC GTG CCG CTA CTC TAC AAG CAA TTT	144
Arg Arg Tyr Ala Arg Lys Gln Leu Val Pro Leu Leu Tyr Lys Gln Phe	
15 20 25	
GTG CCC GGC GTG CCA GAG CGG ACC CTG GGC GCC AGT GGG CCA GCG GAG	192
Val Pro Gly Val Pro Glu Arg Thr Leu Gly Ala Ser Gly Pro Ala Glu	
30 35 40	
GGG AGG GTG GCA AGG GGC TCC GAG CGC TTC CGG GAC CTC GTG CCC AAC	240
Gly Arg Val Ala Arg Gly Ser Glu Arg Phe Arg Asp Leu Val Pro Asn	
45 50 55	
TAC AAC CCC GAC ATC ATC TTC AAG GAT GAG GAG AAC AGT GGA GCC GAC	288
Tyr Asn Pro Asp Ile Ile Phe Lys Asp Glu Asn Ser Gly Ala Asp	
60 65 70	
CGC CTG ATG ACC GAA CGT TGT AAG GAA CGG GTG AAC GCT TTG GCC ATT	336
Arg Leu Met Thr Glu Arg Cys Lys Glu Arg Val Asn Ala Leu Ala Ile	
75 80 85 90	
GCC GTG ATG AAC ATG TGG CCC GGA GTG CGC CTA CGA GTG ACT GAG GGC	384
Ala Val Met Asn Met Trp Pro Gly Val Arg Leu Arg Val Thr Glu Gly	
95 100 105	
TGG GAC GAG GAC GGC CAC CAC GCT CAG GAT TCA CTC CAC TAC GAA GGC	432
Trp Asp Glu Asp Gly His His Ala Gln Asp Ser Leu His Tyr Glu Gly	
110 115 120	
CGT GCT TTG GAC ATC ACT ACG TCT GAC CGC GAC CGC AAC AAG TAT GGG	480
Arg Ala Leu Asp Ile Thr Ser Asp Arg Asp Arg Asn Lys Tyr Gly	
125 130 135	
TTG CTG GCG CGC CTC GCA GTG GAA GCC GGC TTC GAC TGG GTC TAC TAC	528
Leu Leu Ala Arg Leu Ala Val Glu Ala Gly Phe Asp Trp Val Tyr Tyr	
140 145 150	
GAG TCC CGC AAC CAC ATC CAC GTG TCG GTC AAA GCT GAT AAC TCA CTG	576
Glu Ser Arg Asn His Ile His Val Ser Val Lys Ala Asp Asn Ser Leu	
155 160 165 170	
GCG GTC CGG GCG GGC GGC TGC TTT CCG GGA AAT GCA ACT GTG CGC CTG	624
Ala Val Arg Ala Gly Gly Cys Phe Pro Gly Asn Ala Thr Val Arg Leu	
175 180 185	
TGG AGC GGC GAG CGG AAA GGG CTG CGG GAA CTG CAC CGC GGA GAC TGG	672
Trp Ser Gly Glu Arg Lys Gly Leu Arg Glu Leu His Arg Gly Asp Trp	
190 195 200	
GTT TTG ACG GCC GAT GCG TCA GGC CGG GTG GTG CCC ACG CCG GTG CTG	720
Val Leu Thr Ala Asp Ala Ser Gly Arg Val Val Pro Thr Pro Val Leu	
205 210 215	
CTC TTC CTG GAC CGG GAC TTG CAG CGC CGG GCT TCA TTT GTG GCT GTG	768
Leu Phe Leu Asp Arg Asp Leu Gln Arg Arg Ala Ser Phe Val Ala Val	
220 225 230	
GAG ACC GAG TGG CCT CCA CGC AAA CTG TTG CTC ACG CCC TGG CAC CTG	816
Glu Thr Glu Trp Pro Pro Arg Lys Leu Leu Thr Pro Trp His Leu	
235 240 245 250	
GTG TTT GCC GCT CGA GGG CCG GCG CCC GCG CCA GGC GAC TTT GCA CCG	864
Val Phe Ala Ala Arg Gly Pro Ala Pro Ala Pro Gly Asp Phe Ala Pro	
255 260 265	

GTC	TTC	GCG	CGC	CGG	CTA	CGC	GCT	GGG	GAC	TCG	GTG	CTG	GCG	CCC	GGC	912
Val	Phe	Ala	Arg	Arg	Leu	Arg	Ala	Gly	Asp	Ser	Val	Leu	Ala	Pro	Gly	
																270
																275
																280
GGG	GAT	GCG	CTT	CGG	CCA	GCG	CGC	GTG	GCC	CGT	GTG	GCG	CGG	GAG	GAA	960
Gly	Asp	Ala	Leu	Arg	Pro	Ala	Arg	Val	Ala	Arg	Val	Ala	Arg	Glu	Glu	
																285
																290
																295
GCC	GTG	GGC	GTG	TTC	GCG	CCG	CTC	ACC	GCG	CAC	GGG	ACG	CTG	CTG	GTG	1008
Ala	Val	Gly	Val	Phe	Ala	Pro	Leu	Thr	Ala	His	Gly	Thr	Leu	Leu	Val	
																300
																305
																310
AAC	GAT	GTC	CTG	GCC	TCT	TGC	TAC	GCG	GTT	CTG	GAG	AGT	CAC	CAG	TGG	1056
Asn	Asp	Val	Leu	Ala	Ser	Cys	Tyr	Ala	Val	Leu	Glu	Ser	His	Gln	Trp	
																315
																320
																325
																330
GCG	CAC	CGC	GCT	TTT	GCC	CCC	TTG	AGA	CTG	CTG	CAC	GCG	CTA	GGG	GCG	1104
Ala	His	Arg	Ala	Phe	Ala	Pro	Leu	Arg	Leu	Leu	His	Ala	Leu	Gly	Ala	
																335
																340
																345
CTG	CTC	CCC	GGC	GGG	GCC	GTC	CAG	CCG	ACT	GGC	ATG	CAT	TGG	TAC	TCT	1152
Leu	Leu	Pro	Gly	Gly	Ala	Val	Gln	Pro	Thr	Gly	Met	His	Trp	Tyr	Ser	
																350
																355
																360
CGG	CTC	CTC	TAC	CGC	TTA	GCG	GAG	GAG	CTA	CTG	GGC					1188
Arg	Leu	Leu	Tyr	Arg	Leu	Ala	Glu	Glu	Leu	Leu	Gly					
																365
																370

(2) INFORMATION FOR SEQ ID NO: 7:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 548 base pairs
 - (B) TYPE: nucleic acid
 - (C) strandedness: double
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: human
 - (B) INDIVIDUAL ISOLATE: ARH-77, ATCC CRL-1621
- (ix) FEATURE:
 - (A) NAME/KEY: sig peptide
 - (B) LOCATION: 1..18
 - (C) IDENTIFICATION METHOD: S
 - (A) NAME/KEY: mat peptide
 - (B) LOCATION: 19..546
 - (C) IDENTIFICATION METHOD: S

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

GCG	CTG	CCA	GCC	CAG	AGC	TGC	GGG	CCG	GGC	GGG	CCG	GTT	GGC	CGG	48	
Ala	Leu	Pro	Ala	Gln	Ser	Cys	Gly	Pro	Gly	Arg	Gly	Pro	Val	Gly	Arg	
-5						1			5			10				
CGC	CGC	TAT	GCG	CGC	AAG	CAG	CTC	GTG	CCG	CTA	CTC	TAC	AAG	CAA	96	
Arg	Arg	Tyr	Ala	Arg	Lys	Gln	Leu	Val	Pro	Leu	Leu	Tyr	Lys	Gln	Phe	
															15	
															20	
															25	
GTG	CCC	GGC	GTG	CCA	GAG	CGG	ACC	CTG	GGC	GCC	AGT	GGG	CCA	GCG	GAG	144
Val	Pro	Gly	Val	Pro	Glu	Arg	Thr	Leu	Gly	Ala	Ser	Gly	Pro	Ala	Glu	
															30	
															35	
															40	
GGG	AGG	GTG	GCA	AGG	GGC	TCC	GAG	CGC	TTC	CGG	GAC	CTC	GTG	CCC	AAC	192
Gly	Arg	Val	Ala	Arg	Gly	Ser	Glu	Arg	Phe	Arg	Asp	Leu	Val	Pro	Asn	
															45	
															50	
															55	

TAC AAC CCC GAC ATC ATC TTC AAG GAT GAG GAG AAC AGT GGA GCC GAC	240
Tyr Asn Pro Asp Ile Ile Phe Lys Asp Glu Glu Asn Ser Gly Ala Asp	
60 65 70	
CGC CTG ATG ACC GAA CGT TGT AAG GAA CGG GTG AAC GCT TTG GCC ATT	288
Arg Leu Met Thr Glu Arg Cys Lys Glu Arg Val Asn Ala Leu Ala Ile	
75 80 85 90	
GCC GTG ATG AAC ATG TGG CCC GGA GTG CGC CTA CGA GTG ACT GAG GGC	336
Ala Val Met Asn Met Trp Pro Gly Val Arg Leu Arg Val Thr Glu Gly	
95 100 105	
TGG GAC GAG GAC GGC CAC CAC GCT CAG GAT TCA CTC CAC TAC GAA GGC	384
Trp Asp Glu Asp Gly His His Ala Gln Asp Ser Leu His Tyr Glu Gly	
110 115 120	
CGT GCT TTG GAC ATC ACT ACG TCT GAC CGC GAC CGC AAC AAG TAT GGG	432
Arg Ala Leu Asp Ile Thr Thr Ser Asp Arg Asp Arg Asn Lys Tyr Gly	
125 130 135	
TTG CTG GCG CGC CTC GCA GTG GAA GCC GGC TTC GAC TGG GTC TAC TAC	480
Leu Leu Ala Arg Leu Ala Val Glu Ala Gly Phe Asp Trp Val Tyr Tyr	
140 145 150	
GAG TCC CGC AAC CAC ATC CAC GTG TCG GTC AAA GCT GAT AAC TCA CTG	528
Glu Ser Arg Asn His Ile His Val Ser Val Lys Ala Asp Asn Ser Leu	
155 160 165 170	
CGC GTC CGG GCG GGC GGC TG	548
Ala Val Arg Ala Gly Gly	
175	

(2) INFORMATION FOR SEQ ID NO: 8:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 602 base pairs
 - (B) TYPE: nucleic acid
 - (C) strandedness: double
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: human
 - (B) INDIVIDUAL ISOLATE: ARH-77, ATCC CRL-1621
- (ix) FEATURE:
 - (A) NAME/KEY: 5'UTR
 - (B) LOCATION: 1..6
 - (C) IDENTIFICATION METHOD: S
 - (A) NAME/KEY: sig peptide
 - (B) LOCATION: 7..72
 - (C) IDENTIFICATION METHOD: S
 - (A) NAME/KEY: mat peptide
 - (B) LOCATION: 73..600
 - (C) IDENTIFICATION METHOD: S

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

GTATCC ATG GCT CTC CTG ACC AAT CTA CTG CCC CTG TGC TGC TTG GCA	48
Met Ala Leu Leu Thr Asn Leu Leu Pro Leu Cys Cys Leu Ala	
-20 -15 -10	
CTT CTG GCG CTG CCA GCC CAG AGC TGC GGG CCG GGC CGG GGG CCG GTT	96
Leu Leu Ala Leu Pro Ala Gln Ser Cys Gly Pro Gly Arg Gly Pro Val	
-5 1 5	

GGC CGG CGC CGC TAT GCG CGC AAG CAG CTC GTG CCG CTA CTC TAC AAG	144
Gly Arg Arg Arg Tyr Ala Arg Lys Gln Leu Val Pro Leu Leu Tyr Lys	
10 15 20	
CAA TTT GTG CCC GGC GTG CCA GAG CGG ACC CTG GGC GCC AGT GGG CCA	192
Gln Phe Val Pro Gly Val Pro Glu Arg Thr Leu Gly Ala Ser Gly Pro	
25 30 35 40	
GCG GAG GGG AGG GTG GCA AGG GGC TCC GAG CGC TTC CGG GAC CTC GTG	240
Ala Glu Gly Arg Val Ala Arg Gly Ser Glu Arg Phe Arg Asp Leu Val	
45 50 55	
CCC AAC TAC AAC CCC GAC ATC ATC TTC AAG GAT GAG GAG AAC AGT GGA	288
Pro Asn Tyr Asn Pro Asp Ile Ile Phe Lys Asp Glu Glu Asn Ser Gly	
60 65 70	
GCC GAC CGC CTG ATG ACC GAA CGT TGT AAG GAA CGG GTG AAC GCT TTG	336
Ala Asp Arg Leu Met Thr Glu Arg Cys Lys Glu Arg Val Asn Ala Leu	
75 80 85	
GCC ATT GCC GTG ATG AAC ATG TGG CCC GGA GTG CGC CTA CGA GTG ACT	384
Ala Ile Ala Val Met Asn Met Trp Pro Gly Val Arg Leu Arg Val Thr	
90 95 100	
GAG GGC TGG GAC GAG GAC GGC CAC CAC GCT CAG GAT TCA CTC CAC TAC	432
Glu Gly Trp Asp Glu Asp Gly His His Ala Gln Asp Ser Leu His Tyr	
105 110 115 120	
GAA GGC CGT GCT TTG GAC ATC ACT ACG TCT GAC CGC GAC CGC AAC AAG	480
Glu Gly Arg Ala Leu Asp Ile Thr Ser Asp Arg Asp Arg Asn Lys	
125 130 135	
TAT GGG TTG CTG GCG CGC CTC GCA GTG GAA GCC GGC TTC GAC TGG GTC	528
Tyr Gly Leu Leu Ala Arg Leu Ala Val Glu Ala Gly Phe Asp Trp Val	
140 145 150	
TAC TAC GAG TCC CGC AAC CAC ATC CAC GTG TCG GTC AAA GCT GAT AAC	576
Tyr Tyr Glu Ser Arg Asn His Ile His Val Ser Val Lys Ala Asp Asn	
155 160 165	
TCA CTG GCG GTC CGG GCG GGC TG	602
Ser Leu Ala Val Arg Ala Gly Gly	
170 175	

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 575 base pairs
- (B) TYPE: nucleic acid
- (C) strandedness: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: human
- (B) INDIVIDUAL ISOLATE: ARH-77, ATCC CRL-1621

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

C GTG TCG GTC AAA GCT GAT AAC TCA CTG GCG GTC CGG GCG GGC GGC	46
Val Ser Val Lys Ala Asp Asn Ser Leu Ala Val Arg Ala Gly Gly	
1 5 10 15	
TGC TTT CCG GGA AAT GCA ACT GTG CGC CTG TGG AGC GGC GAG CGG AAA	94
Cys Phe Pro Gly Asn Ala Thr Val Arg Leu Trp Ser Gly Glu Arg Lys	
20 25 30	

GGG CTG CGG GAA CTG CAC CGC GGA GAC TGG GTT TTG ACG GCC GAT GCG	142
Gly Leu Arg Glu Leu His Arg Gly Asp Trp Val Leu Thr Ala Asp Ala	
35	40
45	
TCA GGC CGG GTG GTG CCC ACG CCG GTG CTG CTC TTC CTG GAC CGG GAC	190
Ser Gly Arg Val Val Pro Thr Pro Val Leu Leu Phe Leu Asp Arg Asp	
50	55
60	
TTG CAG CGC CGG GCT TCA TTT GTG GCT GTG GAG ACC GAG TGG CCT CCA	238
Leu Gln Arg Arg Ala Ser Phe Val Ala Val Glu Thr Glu Trp Pro Pro	
65	70
75	
CGC AAA CTG TTG CTC ACG CCC TGG CAC CTG GTG TTT GCC GCT CGA GGG	286
Arg Lys Leu Leu Leu Thr Pro Trp His Leu Val Phe Ala Ala Arg Gly	
80	85
90	95
CCG GCG CCC GCG CCA GGC GAC TTT GCA CCG GTG TTC GCG CGC CGG CTA	334
Pro Ala Pro Ala Pro Gly Asp Phe Ala Pro Val Phe Ala Arg Arg Leu	
100	105
110	
CGC GCT GGG GAC TCG GTG CTG GCG CCC GGC GGG GAT GCG CTT CGG CCA	382
Arg Ala Gly Asp Ser Val Leu Ala Pro Gly Gly Asp Ala Leu Arg Pro	
115	120
125	
CGC CGC GTG GCC CGT GTG GCG CGG GAG GAA GCC GTG GGC GTG TTC GCG	430
Ala Arg Val Ala Arg Val Ala Arg Glu Glu Ala Val Gly Val Phe Ala	
130	135
140	
CCG CTC ACC GCG CAC GGG ACG CTG CTG GTG AAC GAT GTC CTG GCC TCT	478
Pro Leu Thr Ala His Gly Thr Leu Leu Val Asn Asp Val Leu Ala Ser	
145	150
155	
TGC TAC GCG GTT CTG GAG AGT CAC CAG TGG GCG CAC CGC GCT TTT GCC	526
Cys Tyr Ala Val Leu Glu Ser His Gln Trp Ala His Arg Ala Phe Ala	
160	165
170	175
CCC TTG AGA CTG CTG CAC GCG CTA GGG GCG CTG CTC CCC GGC GGG GCC	574
Pro Leu Arg Leu Leu His Ala Leu Gly Ala Leu Leu Pro Gly Gly Ala	
180	185
190	
G	575

(2) INFORMATION FOR SEQ ID NO: 10:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 230 base pairs
 - (B) TYPE: nucleic acid
 - (C) strandedness: double
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: human
 - (B) INDIVIDUAL ISOLATE: ARH-77, ATCC CRL-1621
- (ix) FEATURE:
 - (A) NAME/KEY: 3'UTR
 - (B) LOCATION: 218..230
 - (C) IDENTIFICATION METHOD: S
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

G TTC GCG CCG CTC ACC GCG CAC GGG ACG CTG CTG GTG AAC GAT GTC	46
Phe Ala Pro Leu Thr Ala His Gly Thr Leu Leu Val Asn Asp Val	
1	5
10	
15	

CTG	GCC	TCT	TGC	TAC	GCG	GTT	CTG	GAG	AGT	CAC	CAG	TGG	GCG	CAC	CGC		94	
Leu	Ala	Ser	Cys	Tyr	Ala	Val	Leu	Glu	Ser	His	Gln	Trp	Ala	His	Arg			
																30		
20										25								
GCT	TTT	GCC	CCC	TTG	AGA	CTG	CTG	CAC	GCG	CTA	GGG	GCG	CTG	CTC	CCC		142	
Ala	Phe	Ala	Pro	Leu	Arg	Leu	Leu	His	Ala	Leu	Gly	Ala	Leu	Leu	Pro			
																35		
																40		
																	45	
GGC	GGG	GCC	GTC	CAG	CCG	ACT	GGC	ATG	CAT	TGG	TAC	TCT	CGG	CTC	CTC		190	
Gly	Gly	Ala	Val	Gln	Pro	Thr	Gly	Met	His	Trp	Tyr	Ser	Arg	Leu	Leu			
																50		
																55		
																	60	
TAC	CGC	TTA	GCG	GAG	GAG	CTA	CTG	GGC	TGAGCGTCCC	AGG							230	
Tyr	Arg	Leu	Ala	Glu	Glu	Leu	Leu	Gly										
																65		
																70		

(2) INFORMATION FOR SEQ ID NO: 11:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 522 base pairs
 - (B) TYPE: nucleic acid
 - (C) strandedness: double
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: human
 - (B) INDIVIDUAL ISOLATE: A549, ATCC CRL-185
- (ix) FEATURE:
 - (A) NAME/KEY: mat peptide
 - (B) LOCATION: 1..522
 - (C) IDENTIFICATION METHOD: S

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

TGC	GGA	CCG	GGC	AGG	GGG	TTC	GGG	AAA	AGG	AGG	CAC	CCC	AAA	AAG	CTG		48
Cys	Gly	Pro	Gly	Arg	Gly	Phe	Gly	Lys	Arg	Arg	His	Pro	Lys	Lys	Leu		
1						5					10				15		
ACC	CCT	TTA	GCC	TAC	AAG	CAG	TTT	ATC	CCC	AAT	GTG	GCC	GAA	AAG	ACC		96
Thr	Pro	Leu	Ala	Tyr	Lys	Gln	Phe	Ile	Pro	Asn	Val	Ala	Glu	Lys	Thr		
																20	
																25	
																30	
CTA	GGC	GCC	AGC	GGA	AGG	TAT	GAA	GGG	AAG	ATC	TCC	AGA	AAC	TCC	GAG		144
Leu	Gly	Ala	Ser	Gly	Arg	Tyr	Glu	Gly	Lys	Ile	Ser	Arg	Asn	Ser	Glu		
																35	
																40	
																45	
CGA	TTT	AAG	GAA	CTC	ACC	CCC	AAT	TAC	AAC	CCC	GAC	ATC	ATA	TTT	AAG		192
Arg	Phe	Lys	Glu	Leu	Thr	Pro	Asn	Tyr	Asn	Pro	Asp	Ile	Ile	Phe	Lys		
																50	
																55	
																60	
GAT	GAA	GAA	AAC	ACC	GGA	GCG	GAC	AGG	CTG	ATG	ACT	CAG	AGG	TGT	AAG		240
Asp	Glu	Glu	Asn	Thr	Gly	Ala	Asp	Arg	Leu	Met	Thr	Gln	Arg	Cys	Lys		
																65	
																70	
																75	
																80	
GAC	AAG	TTG	AAC	GCT	TTG	GCC	ATC	TCG	GTG	ATG	AAC	CAG	TGG	CCA	GGA		288
Asp	Lys	Leu	Asn	Ala	Leu	Ala	Ile	Ser	Val	Met	Asn	Gln	Trp	Pro	Gly		
																85	
																90	
																95	
GTG	AAA	CTG	CGG	GTG	ACC	GAG	GGC	TGG	GAC	GAA	GAT	GGC	CAC	CAC	TCA		336
Val	Lys	Leu	Arg	Val	Thr	Glu	Gly	Trp	Asp	Glu	Asp	Gly	His	His	Ser		
																100	
																105	
																110	
GAG	GAG	TCT	CTG	CAC	TAC	GAG	GGC	CGC	GCA	GTG	GAC	ATC	ACC	ACG	TCT		384
Glu	Glu	Ser	Leu	His	Tyr	Glu	Gly	Arg	Ala	Val	Asp	Ile	Thr	Thr	Ser		
																115	
																120	
																125	

GAC CGC GAC CGC AGC AAG TAC GGC ATG CTG GCC CGC CTG GCG GTG GAG	432
Asp Arg Asp Arg Ser Lys Tyr Gly Met Leu Ala Arg Leu Ala Val Glu	
130 135 140	
GCC GGC TTC GAC TGG GTG TAC TAC GAG TCC AAG GCA CAT ATC CAC TGC	480
Ala Gly Phe Asp Trp Val Tyr Tyr Glu Ser Lys Ala His Ile His Cys	
145 150 155 160	
TCG GTG AAA GCA GAG AAC TCG GTG GCG GCC AAA TCG GGA GGC	522
Ser Val Lys Ala Glu Asn Ser Val Ala Ala Lys Ser Gly Gly	
165 170 174	

(2) INFORMATION FOR SEQ ID NO: 12:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) strandedness: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

GCCAGGGTGT GAGCAACAGT	20
-----------------------	----

(2) INFORMATION FOR SEQ ID NO: 13:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) strandedness: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

TGTGCTGCTT GGCACTCTTG	20
-----------------------	----

(2) INFORMATION FOR SEQ ID NO: 14:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) strandedness: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

CCGTGGCATT TCCCGGAAAG	20
-----------------------	----

(2) INFORMATION FOR SEQ ID NO: 15:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 18 base pairs
 - (B) TYPE: nucleic acid
 - (C) strandedness: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

GTATCCATGG CTCTCCTG	18
---------------------	----

(2) INFORMATION FOR SEQ ID NO: 16:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 26 base pairs
 - (B) TYPE: nucleic acid
 - (C) strandedness: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

GCCTCGAGGT ATCCATGGCT CTCCTG

26

(2) INFORMATION FOR SEQ ID NO: 17:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) strandedness: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

GCGCGGCCGC TCAGCCGCC CCCCCGGAC

28

(2) INFORMATION FOR SEQ ID NO: 18:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) strandedness: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

CGTGTGGTC AAAGCTGATA

20

(2) INFORMATION FOR SEQ ID NO: 19:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) strandedness: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

ATGCATTCCA GTCGGCTGGA

20

(2) INFORMATION FOR SEQ ID NO: 20:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 56 base pairs
(B) TYPE: nucleic acid
(C) strandedness: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

AAGGATCCGT CGACAAGCTT AATACGACGA ATTCTGGAGT TTTTTTTTTT TTTTTT

56

(2) INFORMATION FOR SEQ ID NO: 21:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) strandedness: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

GGCTTCGACT GGGTCTACTA

20

(2) INFORMATION FOR SEQ ID NO: 22:

- (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 17 base pairs
- (B) TYPE: nucleic acid
- (C) strandedness: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

AAGGATCCGT CGACAAG

17

(2) INFORMATION FOR SEQ ID NO: 23:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 17 base pairs
 - (B) TYPE: nucleic acid
 - (C) strandedness: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

ATGCGCTTCG GCCAGCG

17

(2) INFORMATION FOR SEQ ID NO: 24:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 17 base pairs
 - (B) TYPE: nucleic acid
 - (C) strandedness: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

GACAAGCTTA ATACGAC

17

(2) INFORMATION FOR SEQ ID NO: 25:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 17 base pairs
 - (B) TYPE: nucleic acid
 - (C) strandedness: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

GTTCGCGCCG CTCACCG

17

(2) INFORMATION FOR SEQ ID NO: 26:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 18 base pairs
 - (B) TYPE: nucleic acid
 - (C) strandedness: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

TACGACGAAT TCTGGAGT

18

(2) INFORMATION FOR SEQ ID NO: 27:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 34 base pairs
 - (B) TYPE: nucleic acid
 - (C) strandedness: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

CCCGGGAATT CATTGCGGGC CGGGCCGGGG GCCG

34

(2) INFORMATION FOR SEQ ID NO: 28:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 34 base pairs
 - (B) TYPE: nucleic acid
 - (C) strandedness: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

ACGATGAATT CTCAGCCGCC CGCCCGGACC GCCA

34

(2) INFORMATION FOR SEQ ID NO: 29:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 6 amino acids
 - (B) TYPE: amino acid
 - (C) strandedness: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 29:

Gly Ser Pro Gly Ile His
1 5

(2) INFORMATION FOR SEQ ID NO: 30:

- B/
- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 33 base pairs
 - (B) TYPE: nucleic acid
 - (C) strandedness: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 30:

CCCGGGAATT CATTGCAGGAC CGGGCAGGGG GTT

33

(2) INFORMATION FOR SEQ ID NO: 31:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 32 base pairs
 - (B) TYPE: nucleic acid
 - (C) strandedness: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 31:

ACGATGAATT CTCAGCCTCC CGATTTGGCC GC

32